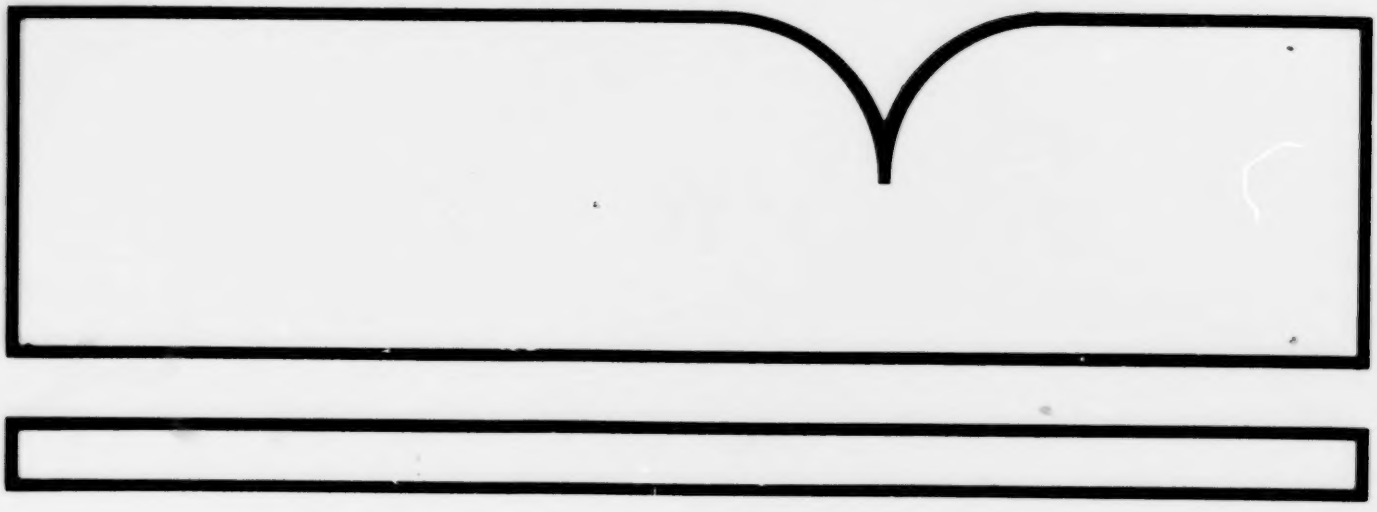
PB83-182931

Evaluation of Automated Spectrum Matching for Survey Identification of Wastewater Components by Gas Chromatography-Mass Spectrometry

(U.S.) Environmental Research Lab.
Athens, GA

25 Jun 82



U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before compl

PB83-182931

1. REPORT NO. EPA-600/J-82-130	2. JOURNAL ARTICLE
4. TITLE AND SUBTITLE Journal Article: An Evaluation of Automated Spectrum Matching for Survey Identification of Wastewater Components by Gas Chromatography--Mass Spectrometry	5. REPORT DATE 1983 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) W.M. Shackelford*, D.M. Cline*, L. Faas** and G. Kurth**	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS *Environmental Research Laboratory Athens, GA **Computer Science Corporation Falls Church, VA	10. PROGRAM ELEMENT NO. ABWD1A 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency--Athens GA Office of Research and Development Environmental Research Laboratory Athens, GA 30613	13. TYPE OF REPORT AND PERIOD COVERED Journal Article 14. SPONSORING AGENCY CODE EPA/600/01

15. SUPPLEMENTARY NOTES
Published in Analytica Chimica Acta, 146(1):15-27, 1983.
Contact: W.M. Shackelford, 8-250-3186

16. ABSTRACT

An automated system for extracting spectra and matching them against a library of reference spectra was assembled and tested on 2000 g.c.-m.s. data files over a period of 2.5 years. The files were for actual field samples and were surveyed for all compounds rather than for a target list. No prior assumptions about sample content could be made. The reliability of spectrum matching is shown to be enhanced from the use of retention data accumulated in an historical library. An overall reliability of 71% is shown for matched spectra in field samples as demonstrated by re-processing of sample extracts.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 14
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (9-73)

EPA-600/J-82-130
JOURNAL ARTICLE

Analytica Chimica Acta, 146 (1983) 15—27
Elsevier Scientific Publishing Company, Amsterdam — Printed in The Netherlands

AN EVALUATION OF AUTOMATED SPECTRUM MATCHING FOR SURVEY IDENTIFICATION OF WASTEWATER COMPONENTS BY GAS CHROMATOGRAPHY—MASS SPECTROMETRY

W. M. SHACKELFORD* and D. M. CLINE

U.S. Environmental Protection Agency, Environmental Research Laboratory, Athens, GA 30613 (U.S.A.)

L. FAAS and G. KURTH

Computer Science Corporation, Falls Church, VA (U.S.A.)

(Received 25th June 1982)

SUMMARY

An automated system for extracting spectra and matching them against a library of reference spectra was assembled and tested on 20000 g.c.-m.s. data files over a period of 2.5 years. The files were for actual field samples and were surveyed for all compounds rather than for a target list. No prior assumptions about sample content could be made. The reliability of spectrum matching is shown to be enhanced from the use of retention data accumulated in an historical library. An overall reliability of 71% is shown for matched spectra in field samples as demonstrated by re-processing of sample extracts.

The use of gas chromatography/mass spectrometry (g.c.-m.s.) with computer-controlled data acquisition and reduction has become the recommended practice for identification of organic pollutants in environmental samples [1, 2]. As can be noted for many samples, however, comparatively few components in any given sample will be a member of some preselected list (such as priority pollutants) [3, 4].

In this study, the probability based matching (PBM) system [5, 6] was used to augment and, in many cases, supplant conventional efforts in component identification so that large numbers of components could be tentatively identified in the most cost-effective way. The goals of this study were: (a) to define the effectiveness of spectrum matching on field samples, and (b) to determine the reliability of the answers generated by spectrum matching. Sample extracts corresponding to selected data files that were processed by the computer were re-examined for the specific compounds identified by spectrum matching [7]. These confirmatory data demonstrate the confidence that can be placed in matching results.

Although previous studies [5—8] have evaluated the use of search systems, the spectra used for unknowns in those studies were known to be of high quality. In this study, the unknown spectra in almost every case were truly

1983 Elsevier Scientific Publishing Company

unknowns in that neither was there prior knowledge of the sample contents available nor was there any guarantee that the spectra submitted for matching were uncontaminated.

Over the 2.5-year duration of this study, some 225 000 spectra were submitted for spectrum matching. Over 30% of the spectra submitted were matched with spectra in the reference library to a degree sufficient that the unknown was considered to be tentatively identified and entered into an historical library.

EXPERIMENTAL

Data used in this study were obtained from a comprehensive study of wastewater from industrial and publicly owned treatment works sponsored by the Effluent Guidelines Division of the U.S. Environmental Protection Agency. Some 4000 samples were taken over a broad range of industrial processes and g.c.-m.s. data were obtained for 114 specific organic compounds (priority pollutants) at twenty different laboratories across the country. All raw g.c.-m.s. data were saved on magnetic tape for survey evaluation. The extract of each extractable fraction from each sample was also sent to the Environmental Research Laboratory, Athens, GA, for confirmation of the survey results. Detailed descriptions of the experimental procedures can be found elsewhere as noted.

Computer system [3]

The computer system consists of four main parts: inventory, spectrum extraction, spectrum matching, and historical library. Because of the number of g.c.-m.s. data files involved (some 22000) a rather elaborate data management system was used for inventory of both the data files and extracts. The INFORM (United Computing Systems, Kansas City, MO) system allows cross-referencing, sorting, and search/retrieval so that correlations of the data by analytical technique, laboratory, or other descriptive parameter are conveniently made.

Spectra were extracted by using the program CLEANUP [9], which was obtained from T. C. Rindfleisch at Stanford University. As reported by the originators of CLEANUP [9], components eluting within two scans of each other cannot, for the most part, be separated. Thus, for complicated data files, many of the extracted spectra are contaminated by one or more other compounds.

Matching an unknown spectrum with a reference spectrum involves two steps, a spectrum match by PBM and a retention data match with the historical library. The spectrum matching program, PBM, was obtained from F. W. McLafferty at Cornell University. A version of PBM evaluated and described previously [5] was used with modifications by Atwater [10] and our group. Modifications such as spectrum subtraction [10], spectrum tilting [10], and reliability ranking [10] were not employed. Both spectrum sub-

traction and spectrum tilting require operator intervention to a greater degree than was desired in this study. Reliability ranking was supplanted by the use of retention data and the use of another ranking factor explained below. The historical library concept has been used in several laboratories [11—15] and was used in this study to add confidence to a given spectrum match by the application of retention data.

Each file examined contained one or more internal standards. Retention data relative to the nearest internal standard were collected for each spectrum chosen for matching. In the acid and base/neutral fractions, only anthracene-d_{10} was used as an internal standard; in the purgeable fractions both bromochloromethane and 1,4-dichlorobutane were used.

To initialize the historical library, several commonly found homologous series were run at two laboratories to generate retention data. Thus, for such series as methyl esters of fatty acids, alkanes, alkylbenzenes, ketones, and others, retention data greatly aided initial identification attempts.

Hardware [3]. For all ADP functions, either a Digital Equipment Corporation PDP 11/70 or PDP 11/34 computer was used. Magnetic discs and tapes were used for both temporary and permanent data storage.

Laboratory procedures [1]

The procedures applied in the twenty laboratories followed a common protocol that has been continuously updated since March 1977 [1]. A purgeable fraction and base/neutral- and acid-extractable fractions for each sample taken in the field were examined by packed column g.c.-m.s. One internal standard, anthracene-d_{10}, was added to the final extract of the acid and base/neutral fraction. Two internal standards, bromochloromethane and 1,4-dichlorobutane, were added to the purgeable fraction. During the whole of the study, two g.c. column liquid phases were used for the purgeable fraction (0.2% Carbowax 1500 and 1% SP-1000), two liquid phases were used for the acid extracts (Tenax GC and 1% SP-1240 DA), and three liquid phases were used for the base/neutral extracts (1% SP-2250, 3% SP-2250, 3% SP-2250 DB). In two sets of data, capillary columns were used for chromatography. One laboratory employed a 0.3-mm i.d., 30-m glass capillary coated with OV-17 for both acid and base/neutral fractions, whereas a second laboratory processed a group of samples using a 0.25-mm i.d., 30-m fused silica capillary coated with SE-54. Column performance was checked by the chromatography of test compounds (benzidine, pentachlorophenol) and g.c.-m.s. performance was required to meet the specifications set forth by Eichelberger et al. [16].

The g.c.-m.s. systems included a Finnigan 3200, Finnigan 4000, Finnigan OWA, Hewlett-Packard 5985, Hewlett-Packard 5982, Hewlett-Packard 5993B, and DuPont Dimaspec. Computers employed in acquisition included Systems Industries System 150, Finnigan Incos, and Hewlett-Packard data systems.

To confirm the tentative identifications made by spectrum matching, selected extracts were chosen and re-processed using glass capillary column g.c.-m.s. at a single laboratory. Because the vast majority of samples had

originally been processed using packed column g.c.-m.s., the expected improvement in chromatographic resolution in the re-examination was deemed a positive influence in determining the presence or absence of a matched component. Capillary columns used in the confirmatory study included Carbowax (0.5-mm i.d., 50-m), SE-30 (0.5-mm i.d., 50-m) and CP SIL-5 (0.25-mm i.d., 25-m).

Confirmation proceeded in two steps. The extract was subjected to capillary column g.c.-m.s. and each of the compounds tentatively identified by spectrum matching was sought in the data from the re-examination. Standards of each of the compounds found in the re-examination were then added to the extract for a second g.c.-m.s. experiment. Confirmation was made on the basis of co-elution with the standard as well as by comparison of the mass spectrum of the unknown with that of the standard.

Reference library [17, 18]

Three reference libraries were evaluated for this study. The Wiley Library of Mass Spectra contains 30476 spectra of 30476 compounds. The EPA-NIH Library contains 34363 compounds, many of which are in the Wiley Library also. Finally, the EPA master data base that contains over 40000 spectra of about 32000 compounds was available locally. Because all three libraries contained spectra unique to each, the libraries were combined.

RESULTS AND DISCUSSION

Matching effectiveness

The success of the matching program can be gauged by the number of correct and incorrect answers returned from a known set of data. In this work, however, the estimation of matching success depends upon several unknown factors which complicate matters somewhat. First, spectra were chosen by a computer algorithm that, although having built-in allowances for noise and column bleed, could not distinguish between many common chromatography artifacts and the compounds of interest. For instance, if an instrument had a poorly conditioned septum, the algorithm found the resulting silicone peaks as if they were sample components. Neither the matching program nor the data analysis could specifically identify spectra of this sort for the most part, although the chemical class of silicones was obvious. Also, if sample components eluting at the end of the temperature program appeared to be broad peaks not unlike column bleed artifacts, the CLEANUP program would ignore them in many cases.

A second complicating factor was that only in a very few cases, even with standards, were all the components separated by the chromatograph to such an extent that pure spectra were obtained for all components. In cases of contamination, the analyst must make the decision whether one or more components are present. This decision, in all but the most obvious of cases (Fig. 1), was found to be difficult.

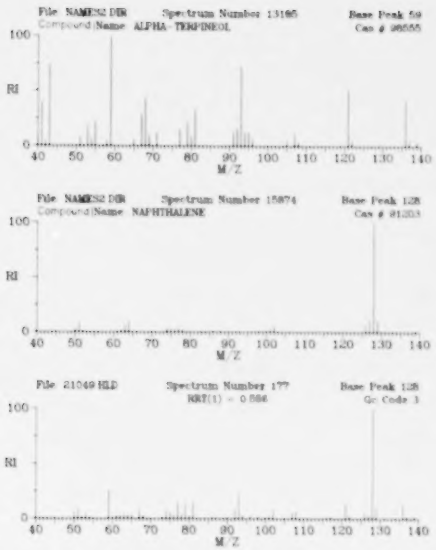


Fig. 1. Spectrum of two components not separated by CLEANUP but both identified by PBM.

Finally, in this study the matches were reconfirmed by capillary g.c.-m.s. In no case was there any guarantee that components in the extract would not degrade over time. In fact, degradation of some components was seen to occur at the confirming laboratory during the time period between the initial and confirmatory studies.

Table 1 summarizes, by gas chromatographic column and fraction type, the overall effectiveness of the automated system. It is important to note that all spectra sent to the matching program were chosen by the CLEANUP algorithm. Thus, even though there are errors of omission and commission in the choosing of spectra, these errors should be constant across the various

TABLE 1

Comparison of g.c. columns and fraction type with matching efficiency

Column	Fraction type	Spectra matched (%)	
		Samples	Standards
0.2% Carbowax 1500 on Carbopak C (packed)	Purgeable	57	70
SP-1240 DA (packed)	Acid	27	77
3% SP-2250 (packed)	Base/neutral	26	71
SE-54 fused silica (capillary)	Acid-base/ neutral (combined)	65	85

data sets. The information to be gained from Table 1 is twofold. First, it is apparent from the data shown that the capillary column chromatography results in a greater fraction of matched spectra than with the packed columns. Second, the compounds present in the fraction make a decided difference in the matching efficiency. Note that the purgeable fraction, though chromatographed on a packed column, yielded 57% matchable spectra. It is believed that this efficiency is due to the preponderance of compounds with distinctive spectra such as low-molecular-weight halogenated compounds. For most samples, the base/neutral and acid fractions contained a large number of obviously unresolved components. Because the possible set of answers was in no way thought to be a subset of the reference library, discrimination to false positive results by PBM was tested severely.

The data on standards showed very little difference among the matching efficiencies for the packed columns. The acid fraction contained only 11 compounds, whereas the purgeable and base/neutral fractions contained 31 and 42 compounds respectively. With standards also, the capillary column showed an appreciable advantage in providing matchable spectra.

The data in Table 1 should be viewed not as an absolute indication of spectral matching ability, but rather as a relative measure of experimental factors that improve matching ability. For instance, if the possible set of sample components is limited, as in the purgeable fraction, the rate of component matching will be higher. Likewise, the better the chromatographic resolution, the better will be the matching efficiency. The values in this table reflect the leniency of component recognition by CLEANUP as much as the matching ability of PBM. In order to insure that components were not missed by CLEANUP, spectra were accepted that turned out to be artifacts.

Extracts containing compounds found frequently by the matching system were sent to the confirming laboratory for study. There were 827 compounds to be confirmed in the 436 extracts sent for study. Because many of the compounds were found in more than one extract, each occurrence of a compound was confirmed. Of the total of 2373 occurrences to be confirmed, 71% were found in the initial re-examination by capillary column g.c.-m.s. The occurrence of the new results with PBM at this point corresponds to matching at the Class II level as defined by Pesyna et al. [5] (identical compound, stereoisomer, or positional isomer). Of those confirmed at the Class II level, 91% were additionally confirmed at the Class I level (identical compound or stereoisomer) by co-injection of standards.

The extracts used for the confirmation had been stored for periods of time up to 2.5 years, and no data exist for the great majority of compounds for stability in dichloromethane extracts. Although a concerted effort was made to insure correct documentation for each sample through quality control procedures, it is apparent that errors are present that cause wrong assignment of an extract to the wrong g.c.-m.s. data file. Because about 8% of the extracts re-examined showed no correspondence with either the PBM results or the reconstructed gas chromatogram of the original sample, we estimate this to be the upper limit of documentation errors.

TABLE 2

Comparison of retention data with K and K_{max} for selected compounds

Compound (CASRN)	Number of occurrences	G.c. column	S.d.[a]	K		K/K_{max}	
				Median	Range	Median	Range
2-Butanone [78-99-3]	67	0.2% Carbowax 1500	±0.062	36	62–20	0.57	0.86–0.41
1,1,2,2-Tetrachloroethane [79-34-5]	93	0.2% Carbowax 1500	±0.007	65	99–25	0.65	1.00–0.26
α-Pinene [80-56-8]	82	0.2% Carbowax 1500	±0.067	67	89–42	0.74	1.00–0.48
2-Chlorophenol [95-57-8]	18	SP-1240 DA	±0.043	52	78–38	0.56	0.79–0.41
Benzoic acid [65-85-0]	298	SP-1240 DA	±0.035	66	100–22	0.74	1.00–0.38
Stearic acid [57-11-4]	66	SP-1240 DA	±0.053	88	138–44	0.63	0.95–0.30
α-Terpineol [98-55-5]	224	3% SP-2250	±0.041	72	97–38	0.74	1.00–0.38
Diethylphthalate [84-66-2]	222	3% SP-2250	±0.012	75	113–36	0.66	1.00–0.32
Cholesterol [57-88-6]	79	3% SP-2250	±0.055	111	193–41	0.54	1.00–0.19
α-Picoline [108-89-4]	21	SE-54 FSCC	±0.007	59	74–43	0.64	0.80–0.46
Nitrobenzene [98-95-3]	20	SE-54 FSCC	±0.004	73	83–45	0.82	0.89–0.47
Fluorenone [486-25-9]	10	SE-54 FSCC	±0.001	87	95–55	0.92	1.00–0.58

[a]Standard deviation of the relative retention.

The 9% of the identified target compounds that was not confirmed by co-injection of standards was found to be more or less equally divided between errors in determining the correct isomer and errors in identification by PBM that were also made by manual identification.

Although it is difficult to assess quantitatively the effectiveness of the matching system because of the unknown variables in this work, it is safe to say that a minimum of 65% (91% of 71%) of those matches accepted by the system were correct even for specific isomers. As will be demonstrated below, this is a greater overall reliability than has been predicted in other work [5, 10] for similar match qualities.

Match quality

Table 2 illustrates, for selected compounds on four types of chromatographic column, the range of match quality parameters (K and K/K_{max}) that were accepted for tentative identification purposes. As can be seen, the range of K values was in many cases greater than a factor of two. Likewise, the range of K/K_{max} values is as great as a factor of five. Examination of the standard deviation of the relative retention, however, shows that the retention data in this work are defined within rather narrow limits. It is the retention data coupled with the match quality that lead to efficient matching. As demonstrated in the retention data column, the fused silica capillary column data show that much tighter retention data windows are possible with capillary column chromatography for use in corroboration of matches.

Use of relative retention times improves matching efficiency, also, in that as more corroborating evidence is acquired, a greater number of tentative identifications can be made with the same spectral information. Table 3 illustrates the increase in matching efficiency (tentative identifications/spectra matched) as more relative retention times were entered in the historical library.

Figure 2 demonstrates the effect of retention data corroboration on the confidence that can be placed in the K value for match quality. The data of Atwater [10] were used to show the reliability of K in distinguishing the correct match for a given spectrum. Plots of data from this study show much higher reliability at lower K values than those of Atwater [10]. This higher reliability is due to the use of retention data as part of the matching process.

TABLE 3

Increase of matching efficiency with the growth of the historical library

Number of compounds in historical library	Date	Matching efficiency (%)
620	December 79	18
945	July 80	23
1553	September 81	31

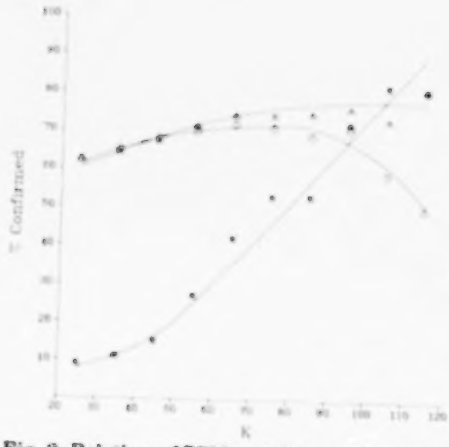


Fig. 2. Relation of PBM matches confirmed to overall match quality K. (0) All compounds; (△) carboxylic acids deleted; (•) data of Atwater [10].

At $K > 95$, however, an anomaly was seen. Instead of the reliability increasing, it began to decrease, unlike the finding of Atwater [10]. Examination of the matches with high K values for compounds that had been sent for confirmation showed that a large fraction of the unconfirmed answers were carboxylic acids. Further examination showed that, for fatty acids of C_{11} or greater, only 32% were confirmed although most had K values in excess of 100. Likewise, aromatic carboxylic acids had only a 37% confirmation rate. All results for carboxylic acids were then removed from consideration and the data were plotted. As can be seen in Fig. 2, the remaining data follow closely the data of Atwater [10] at $K > 95$.

The reason for the poor confirmation rate for the carboxylic acid spectra is not known, although it is believed that sample degradation is in some measure responsible. In Fig. 3 the relation of match reliability to compound class is shown. Whereas one would expect the ambiguity of matching results among alkanes and alkylbenzenes to cause poor reliability, it appears that stability of the compounds during storage has a great influence on reliability in this work. Esters of carboxylic acids had a much greater reliability in matching than the acids themselves.

In Fig. 4 the relation of reliability to ΔK is shown. At low ΔK, the reliabilities are similar to those of Atwater [10] and at high ΔK matches in this work have much higher reliabilities because of the retention data corroboration. The ΔK plot did not show the anomalous decrease at low ΔK corresponding to that at high K seen in Fig. 2. The reason is that the ΔK value is a measure of the difference between the unknown and the reference and as such is not as greatly influenced by the molecular weight as K is. For instance, as molecular weight increases, the highest possible K value increases generally because there will be more peaks in the spectrum. The calculated K value

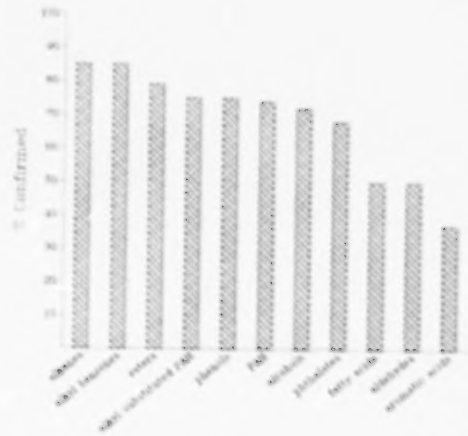


Fig. 3. Relation of PBM matches confirmed to compound class.

may be larger than 100, denoting an excellent match, and yet ΔK can be large also, denoting some question in the match. The set of compounds that predominated in the $K > 95$ regions did not, moreover, predominate in the $\Delta K \leq 15$ region. The carboxylic acids of high molecular weight (>184) that showed such a poor confirmation rate were evenly distributed with ΔK values up to 100, even though their K values had been concentrated in the >95 categories.

Observations like those above led to the use of K/K_{max} as the ranking value rather than K alone. In this way, high K values with large ΔK values

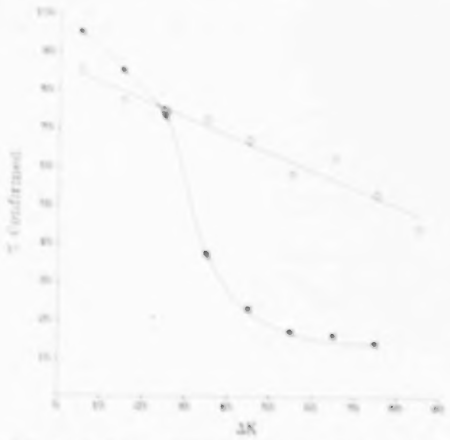


Fig. 4. Relation of PBM matches confirmed to match deficiency, ΔK. (○) All compounds; (•) data of Atwater.

were not given undue weight over small K values with small ΔK values. In Fig. 5, the relation of K/K_{max} to the rate of confirmation at constant K ranges is displayed. The positive slope of each line indicates the utility of K/K_{max} as a ranking factor.

Spectrum quality

For the most part, the reference spectra available for matching are of known quality [19]. The quality of spectra input for matching, however, is generally unknown when submitted in automated fashion. In order to make some measure of the quality of the spectra that had been tentatively identified by the matching software, a quality index [19] was calculated for approximately 25% of the matched spectra. In Table 4 results of applying the quality index calculations to the above set of spectra are compared to those of a previous study by Speck et al. [19]. As can be seen, spectra matched in this work appear to be of lower overall quality than those in reference libraries. If Table 4 is examined closely, however, one can see that the individual quality factors (QF1—QF7) compare quite well except for QF6. In fact, with the exception of this quality factor, the spectra of this study show quality factors of equal or better value than the reference libraries.

The factor, QF6, is a measure of the number of peaks present in the spectrum compared to the number of peaks expected from the molecular formula. Apparently, the spectrum extraction program was unable to separate the small peaks of the spectra from the background on the average. Such points as isotopic ratio accuracy (QF5), high-molecular-weight impurities (QF3), and illogical neutral losses (QF4) all appear to be well within reference

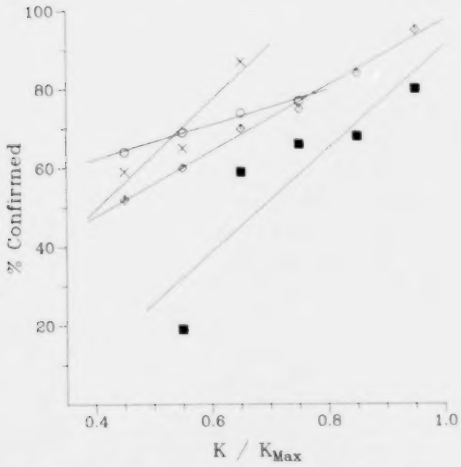


Fig. 5. Plot of K/K_{max} versus confirmation rate at constant K ranges. (×) K = 20—40; (○) K = 41—60; (◇) K = 61—80; (■) K > 80.

TABLE 4

Comparison of quality indices and quality factors for spectra from several sources

Quality factor	Reference library A	Reference library B	Reference library C	Literature [19]	Present work
QF1	1.0(0.0)[a]	0.909(0.0)	0.900(0.0)	0.927(0.0)	0.900(0.0)
QF2	0.995(0.65)	0.962(4.5)	0.972(4.1)	0.999(0.05)	1.000(0.0)
QF3	0.988(0.95)	0.976(1.9)	0.990(0.89)	0.985(1.4)	0.954(3.2)
QF4	0.979(1.1)	0.964(1.9)	0.976(1.5)	0.975(1.8)	0.970(1.3)
QF5	0.965(0.65)	0.971(0.56)	0.955(1.2)	0.942(1.5)	0.977(0.45)
QF6	0.848(13.7)	0.873(10.3)	0.873(10.6)	0.748(24.9)	0.671(28.2)
QF7	0.931(3.8)	0.968(1.4)	0.953(1.8)	0.886(6.9)	0.926(0.23)
QI[b]	0.756(21.2)	0.689(22.6)	0.691(22.4)	0.579(40.3)	0.505(51.0)

[a] Values in parentheses represent percent of spectra showing QF < 0.5. [b] $QI = \Pi_{N=1}^{7} QF_N$.

spectrum quality. Thus, the spectrum extraction program appears to have delivered accurate spectra even though the threshold for peak detection seems to have been high in some cases.

Conclusions

The use of an automated spectrum matching system for survey identification by g.c.-m.s. is quite feasible with little intervention. The reliability of matching when an historical library of retention data is used in conjunction with the spectrum library is at least 65% including matching specific isomers. In this study, 71% overall reliability was achieved when specific isomer resolution was not required. Use of retention data greatly enhances the reliability of lower quality matches, but has little effect on high quality matches. Both positive and negative points of the match must be considered before reliability is assigned.

Valuable discussions with Professors F. W. McLafferty and J. McGuire are gratefully acknowledged. The assistance of Barbara Atwater Fell in bringing PBM up on our PDP 11/70 system is appreciated as is the help of In Ki Mun and David Martinsen with the quality index program and Drew Sauter with the assembly of the total package. We thank Bruce Bartel and Susan Sims for generating the statistical reports necessary for this work, and Laura Folwell for her help in collation of results.

REFERENCES

1 Federal Register, 44 (233) 69532–69552, December 3, 1980.
2 L. H. Keith and W. A. Telliard, Environ. Sci. Technol., 13 (1979) 416.
3 W. M. Shackelford, D. M. Cline, L. O. Burchfield, L. Faas, G. R. Kurth and A. D. Sauter, in L. H. Keith (Ed.), Advances in the Identification and Analysis of Organic Pollutants in Water, Ann Arbor Science Publishers, Ann Arbor, 1981.

4 W. M. Shackelford and L. H. Keith, Frequency of Organic Compounds in Water, EPA-600/4-76-062, December 1976.
5 G. M. Pesyna, R. Venkataraghavan, H. E. Dayringer and F. W. McLafferty, Anal. Chem., 48 (1976) 1362.
6 F. W. McLafferty, R. H. Hertil and R. D. Villwock, Org. Mass Spectrom., 9 (1974) 690.
7 F. P. Abramson, Anal. Chem., 47 (1975) 45.
8 G. T. Rasmussen and T. L. Isenhour, J. Chem. Inf. Comput. Sci., 19 (1979) 179.
9 R. G. Dromey, M. J. Stefik, T. C. Rindfleisch and A. M. Duffield, Anal. Chem., 48 (1976) 1368.
10 B. L. Atwater, Ph.D. Thesis, Cornell University, Ithaca, NY, 1980.
11 H. Nau and K. Biemann, Anal. Chem., 46 (1974) 426.
12 C. C. Sweely, N. D. Young, J. F. Holland and S. C. Gates, J. Chromatogr., 99 (1974) 507.
13 B. E. Blaisdell, Anal. Chem., 49 (1977) 180.
14 D. S. Smith, M. Achenback, W. J. Yaeger, P. J. Anderson, W. L. Fitch and T. C. Rindfleisch, Anal. Chem., 49 (1979) 1623.
15 S. Lewis, C. N. Kenyon, J. Meili and A. L. Burlingame, Anal. Chem., 51 (1979) 1275.
16 J. E. Eichelberger, L. E. Harris and W. R. Budde, Anal. Chem., 47 (1975) 995.
17 Registry of Mass Spectral Data (Enlarged Version), Wiley, New York, 1981.
18 EPA/NIH Mass Spectral Data Base, NSRDS-NBS 63.
19 D. D. Speck, R. Venkataraghavan and F. W. McLafferty, Org. Mass Spectrom., 13 (1978) 209.